U.S. PATENT APPLICATION

Inventor(s): Gal MOR

Invention: INTERCONNECT AND GATEWAY PROTECTION IN BIDIRECTIONAL
RING NETWORKS

NIXON & VANDERHYE P.C.
ATTORNEYS AT LAW
1100 NORTH GLEBE ROAD
8TH FLOOR
ARLINGTON, VIRGINIA 22201-4714
(703) 816-4000
Facsimile (703) 816-4100

SPECIFICATION

INTERCONNECT AND GATEWAY PROTECTION IN BIDIRECTIONAL RING NETWORKS

FIELD OF THE INVENTION

The present invention relates generally to communication networks, and specifically to high-speed packet rings.

BACKGROUND OF THE INVENTION

Network ring topologies are gaining in popularity, particularly in Internet Protocol (IP) networks. Such networks enable carriers to offer large bandwidth to users in a cost-effective manner. They also lend themselves to fast rerouting in the event of network failures, since two alternative routes – in clockwise and counterclockwise directions – are generally available for connecting any two nodes on the ring. A drawback of traditional ring implementations, such as SONET/SDH, is that one of the directions is designated as the active ring, while the other direction remains on standby for fault protection when needed. In other words, at any given time, all of the nodes in the ring transmit and receive data only in the active direction. Therefore, ordinarily half of the available bandwidth in these rings is reserved for fault protection and is not exploited under normal operating conditions.

Some recently-developed bidirectional protocols provide more efficient bandwidth utilization by enabling data to be transferred between any pair of nodes in either direction around the ring, while maintaining fast protection against faults. The two opposing traffic directions are commonly referred to as an inner ring and

an outer ring. It will be understood, however, that in the context of the present patent application and in the claims, the terms "inner" and "outer," as well as "clockwise" and "counterclockwise," are used arbitrarily to distinguish between the two opposing directions of packet flow in a ring network. These terms are chosen solely for convenience of explanation, and do not necessarily bear any relation to the physical characteristics of the network.

The leading bidirectional protocol for high-speed packet rings is the Resilient Packet Rings (RPR) protocol, which is in the process of being defined as IEEE standard 802.17. Network-layer routing over RPR is described, for example, by Jogalekar et al., in "IP over Resilient Packet Rings" (Internet Draft draft-jogalekar-iporpr-00), and by Herrera et al., in "A Framework for IP over Packet Transport Rings" (Internet Draft draft-ietf-ipoptr-framework-00). A proposed solution for Media Access Control (MAC - protocol layer 2) in bidirectional ring networks is the Spatial Reuse Protocol (SRP), which is described by Tsiang et al., in Request for Comments (RFC) 2892 of the Internet Engineering Task Force (IETF). These documents are incorporated herein by reference. They are available at www.ietf.org. Using protocols such as these, each node in a ring network can communicate directly with all other nodes through either the inner or the outer ring, using the appropriate Media Access Control (MAC) addresses of the nodes.

Neighboring ring networks are commonly interconnected through one or more interconnect nodes.

Similarly, an interconnect node on a ring network may connect the ring to another network having a different topology. When the interconnect node links a spur or access ring to a main trunk network, the interconnect node is also referred to as a gateway. In cases in which the interconnect node links two ring networks, the node must typically support four times the bandwidth of a single (clockwise or counterclockwise) ring.

When an interconnect node on a ring network fails, the ring network can become isolated from the other networks to which it is normally connected. Even if multiple interconnect nodes are used to connect two networks, a failure in one of the nodes will affect the connections between the networks that previously passed through the failed node. This problem can be solved by installing an additional node as backup to each interconnecting node, and activating the backup node when the node that is normally active fails. In other words, two nodes are installed at the same site, at substantial added cost and demanding additional space, power and maintenance. Duplication of the node also requires that the active and standby nodes follow a mirroring protocol, so that the standby node will be ready to take over when needed. Because each of the two nodes must be able to carry the full bandwidth of the network interconnection that it supports, the backup node cannot practically be used for load sharing. Therefore, the resources of the backup node are wasted most of the time.

SUMMARY OF THE INVENTION

It is an object of some aspects of the present invention to provide improved communication methods and devices for use in bidirectional ring networks.

It is a further object of some aspects of the present invention to provide interconnect nodes for use in bidirectional ring networks that enable rapid resumption of traffic flow in the case of a failure.

It is a further object of some aspects of the present invention to reduce the levels of hardware complication and cost of interconnect nodes in a bidirectional ring network.

In preferred embodiments of the present invention, a bidirectional ring network comprises an interconnect node that links the ring to another network. The other network may or may not be a ring network, as well. The interconnect node comprises a pair of interconnect modules, both of which are coupled to convey outgoing data packets from the ring to the other network and to convey incoming data packets from the other network to the ring network. Normally, as long as both of the modules are operational, the modules share the load of incoming and outgoing traffic, so that interface bandwidth of each of the modules need not be substantially greater than the maximum data rate of one of the two directional rings in the ring network. This load sharing is preferably accomplished by configuring one of the modules to serve only the clockwise ring, while the other module is configured to serve only the counterclockwise ring. Assuming that the other network to which the interconnect node is linked is also a bidirectional ring network, the same coupling scheme and

4

bandwidth limitations apply to the module connections on both of the networks.

The modules communicate with one another, typically via a protection interface, so that each module can detect failures when they occur in the other module. Preferably, each of the modules is preconfigured with the Media Access Control (MAC) address of the other module, as well as its own MAC address. When one of the modules detects a failure on the protection interface, it immediately begins capturing outgoing packets addressed to the other module. In other words, for as long as the failure persists, the non-failing module accepts outgoing packets on both the clockwise and counterclockwise rings. Upon the occurrence of such a failure, however, the nodes in the network preferably redirect the traffic on the rings in order to circumvent the failed module (typically by wrapping or steering the traffic), as provided by bidirectional ring protocols known in the art. Therefore, the total bandwidth of traffic that the non-failing module must handle is still no greater than the maximum data rate of a single ring. Alternatively, the network operator should ensure that the total volume of guaranteed (high-priority) traffic on both rings through the non-failing module will not exceed the capacity of that module. Other low-priority traffic may be delayed or discarded during the protection time, unless the higher-priority services are not utilizing their allocated bandwidth.

Thus, preferred embodiments of the present invention provide a rapid, effective protection scheme for gateways and other interconnect nodes in bidirectional ring networks, while reducing substantially the bandwidth

requirements of such nodes relative to solutions known in the art. Although these preferred embodiments are described with reference to packet networks, and particularly RPR-based networks, it will be appreciated that the principles of the present invention to high-speed data networks of other types.

There is therefore provided, in accordance with a preferred embodiment of the present invention, a communication device for interconnection of first and second networks, of which at least the first network is a ring network over which traffic is transmitted in both clockwise and counterclockwise directions, the device including:

first and second interconnect modules, each such module adapted to receive outgoing data traffic on the first network and to convey the outgoing data traffic to the second network, and to receive incoming data traffic from the second network and to convey the incoming data traffic to the first network,

the first and second interconnect modules being in mutual communication so that while both of the modules are operational, the first module is configured to receive from the first network only the outgoing data traffic transmitted in the clockwise direction, while the second module is configured to receive from the first network only the outgoing data traffic transmitted in the counterclockwise direction, and so that when a fault occurs in one of the first and second modules, the other module receives a fault indication, causing the other module to receive substantially all of the outgoing data traffic and to convey the outgoing data traffic to the second network.

Preferably, the first network is configured to carry the data traffic at a predetermined maximum data rate in each of the clockwise and counterclockwise directions, and wherein each of the modules is interfaced to receive the outgoing data traffic at a data rate not substantially greater than the predetermined maximum. Further preferably, both of the first and second modules are configured to transmit the incoming data traffic over the first network in both the clockwise and counterclockwise directions while both of the modules are operational, as well as when the fault occurs.

In a preferred embodiment, when the fault occurs in one of the modules, the other module is adapted to receive the outgoing data traffic transmitted in the both clockwise and the counterclockwise directions. Preferably, the fault occurs in one of the modules, the other module is adapted to wrap the traffic between the clockwise and counterclockwise directions, away from the module in which the fault occurred.

Preferably, each of the first and second modules is configured with Media Access Control (MAC) addresses assigned to both the first and second modules, and is controlled so that while both of the modules are operational, each of the modules receives only the outgoing data traffic destined to its own MAC address, and so that when the fault occurs in one of the modules, the other module receives the outgoing data traffic destined to both of the MAC addresses.

In a preferred embodiment, the first network includes an access ring network, while the second network includes a trunk network, and the first and second modules are configured to serve as a gateway node between

7

the access ring network and the trunk network. In
another preferred embodiment, the first and second
networks respectively include first and second ring
networks, and the interconnect modules are configured to
convey and receive the traffic to and from the second
network in a manner substantially similar to that in
which the modules convey and receive the traffic to and
from the first network. Alternatively, the second
network is not a ring network.

There is also provided, in accordance with a
preferred embodiment of the present invention, a
communication device for interconnection of first and
second networks, of which at least the first network is a
ring network over which traffic is transmitted in both
clockwise and counterclockwise directions, with a
predetermined maximum data rate in each of the clockwise
and counterclockwise directions, the device including:

first and second interconnect modules configured
with Media Access Control (MAC) addresses assigned
respectively to both the first and second modules, each
such module being adapted to receive outgoing data
traffic on the first network at a data rate not
substantially greater than the predetermined maximum and
to convey the outgoing data traffic to the second
network, and to receive incoming data traffic from the
second network and to convey the incoming data traffic to
the first network,

the first and second interconnect modules being in
mutual communication so that while both of the modules
are operational, each of the modules receives only the
outgoing data traffic destined to its own MAC address,
and so that when a fault occurs in one of the first and

second modules, the other module receives a fault indication, causing the other module to receive the outgoing data traffic destined to both of the MAC addresses and to convey the outgoing data traffic to the second network.

There is additionally provided, in accordance with a preferred embodiment of the present invention, a method for communication between first and second networks, of which at least the first network is a ring network over which traffic is transmitted in both clockwise and counterclockwise directions, the method including:

coupling the first and second networks together via first and second interconnect modules, such that while both of the modules are operational, the first module receives outgoing traffic from the first network for transmission over the second network only in the clockwise direction on the first network, while the second module receives outgoing traffic from the first network for transmission over the second network only in the counterclockwise direction on the first network; and

upon occurrence of a fault in one of the first and second modules, reconfiguring the other of the modules so as to receive substantially all of the outgoing traffic for transmission over the second network.

There is further provided, in accordance with a preferred embodiment of the present invention, a method for communication between first and second networks, of which at least the first network is a ring network over which traffic is transmitted in both clockwise and counterclockwise directions, with a predetermined maximum data rate in each of the clockwise and counterclockwise directions, the method including:

9

coupling the first and second networks together via first and second interconnect modules;

assigning respective Media Access Control (MAC) addresses to the first and second modules and configuring each of the first and second modules with the MAC addresses of both of the modules, such that while both of the modules are operational, each of the modules receives outgoing data traffic on the first network destined to its own MAC address at a data rate not substantially greater than the predetermined maximum and conveys the outgoing data traffic to the second network; and

upon occurrence of a fault in one of the first and second modules, setting the other module to receive the outgoing traffic destined to both of the MAC addresses and to convey the outgoing data traffic to the second network.

The present invention will be more fully understood from the following detailed description of the preferred embodiments thereof, taken together with the drawings in which:

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block diagram that schematically illustrates a system of ring networks, in accordance with a preferred embodiment of the present invention;

Fig. 2 is a block diagram that schematically illustrates a gateway node in a ring network, in accordance with a preferred embodiment of the present invention;

Figs. 3A and 3B are block diagrams that schematically show details of the operation of the

42165S2

gateway node of Fig. 2, in accordance with a preferred
embodiment of the present invention.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

Fig. 1 is a block diagram that schematically shows a network communication system 20, in accordance with a preferred embodiment of the present invention. System 20 comprises an access (or "spur") ring network 22, which is coupled to a trunk (or "metro") ring network 24 by a spur gateway 30. Network 24 is in turn coupled to another network 26, which is not necessarily a ring network, by an interconnect gateway 40. Gateways 30 and 40 are also referred to herein as interconnect nodes. The topology of system 20 is shown here by way of example, for illustrating the operation of gateways 30 and 40, and the extension of the principles described herein to other topologies will be apparent to those skilled in the art.

In addition to gateways 30 and 40, networks 22 and 24 typically comprise nodes 28. The nodes in each network are mutually connected by bidirectional communication media, such as optical fibers or conductive wires. The nodes typically comprise switching equipment, and serve as either access points or gateways to other networks (aggregation points). The communication media in network 22 are configured to define an inner ring 32, over which packets are conveyed between the nodes in a clockwise direction, and an outer ring 34, over which the packets are conveyed in a counterclockwise direction. Similarly, network 24 comprises an inner ring 36 and an outer ring 38. Preferably, both networks 22 and 24 operate in accordance with the RPR protocol described in the Background of the Invention, or with another, similar bidirectional protocol. As noted above, the designations of "inner," "outer," "clockwise" and "counterclockwise" are

12

arbitrary and are used here simply for convenience and clarity of explanation.

Fig. 2 is a block diagram that schematically illustrates details of gateway 30, in accordance with a preferred embodiment of the present invention. Gateway 30 comprises two modules 50 and 52, which are preferably substantially identical. Each module comprises a RPR block 54, connected to transmit and receive data packets over both of rings 32 and 34. The RPR block in module 50 is marked as block 54a, while the RPR block in module 52 is marked as block 54b. A similar numbering scheme is used to refer to the other blocks in modules 50 and 52, as described hereinbelow.

Block 54 in each module is responsible for ring management and performs the MAC-layer functions of capturing packets that are addressed to gateway 30, while passing all other traffic through to the next node along the ring. Each of modules 50 and 52 typically has its own MAC address to which packets may be sent by nodes 28 on both ring 32 and ring 34. Preferably, blocks 54a and 54b are both preconfigured with the MAC addresses of both of the modules. In normal operation, only one of the MAC addresses is active in each of the blocks, and network 22 is preferably configured so that block 54a captures outgoing packets (destined for network 24) only on ring 32, while block 54b captures outgoing packets only on ring 34. Alternatively, the operator of the network may balance the outgoing traffic in other ways, so that both blocks 54a and 54b capture certain outgoing traffic flows on both of the rings. In either case, both blocks 54a and 54b can preferably transmit incoming data packets (received from network 24) on both of rings 32 and 34.

When one of RPR blocks 54 captures a packet addressed to gateway 30, it delivers the packet to a respective traffic processing block 56. This block is typically implemented as a network processor chip that is able to access higher-layer protocol headers at wire speed (to avoid bottlenecks). It is responsible for network-layer functions, such as IP processing, and optionally other higher-level functions, as well, such as Quality of Service (QoS) and network security. Since each of blocks 54 captures packets only on a single one of rings 32 and 34, the bandwidth of blocks 54 and 56 and of the interface between them (at least for outgoing traffic to network 24) need not be any greater than the bandwidth of one of the rings. It will be shown below that this bandwidth is sufficient not only for normal operation, during which both of modules 50 and 52 are active, but also for operation under failure conditions, in which one of the modules serves as a backup for the other one.

Traffic processing block 56 passes outgoing packets from network 22 to a second RPR block 58 on network 24. The functions and interaction of RPR blocks 58a and 58b, with respect to receiving and transmitting packets on inner ring 36 and outer ring 38 of network 24, are substantially similar to those of RPR blocks 54a and 54b on network 22.

A controller 60, typically a suitable microprocessor, monitors and controls the operation of blocks 54, 56 and 58 in each of modules 50 and 52. Controllers 60a and 60b in the two modules exchange signals via a protection interface 62. Normally, the controllers signal one another regularly to indicate that

the respective modules are functioning properly. When a fault occurs in one of the modules, the respective controller may signal the other controller via interface 62 to indicate that the module at fault cannot handle packet traffic, so that the other module should take over receiving the traffic on both rings. Alternatively, one of controllers 60 may sense that there is no response from the other controller, possibly indicating a total failure of the other module, or one of RPR blocks 54a and 54b may receive a failure indication from the other block. Other methods of failure detection will be apparent to those skilled in the art. In all of these cases, the functioning module will take over receiving the traffic destined for the failed module.

The internal structure of gateway 30 is shown in Fig. 2 for the purpose of exemplifying the operation of the present invention in connecting ring networks 22 and 24. It will be understood that the methods of the present invention may be implemented using other architectures, whether for interconnecting different ring networks as in the present example or for interconnecting a ring network with another network of different topology. For instance, gateway 40, coupling networks 24 and 26, preferably comprises a pair of modules (not shown) with RPR blocks on network 24, similar to blocks 54 shown in Fig. 2. On the side connecting to network 26, however, in place of RPR blocks 56, the modules of gateway 40 will have MAC blocks appropriate to the structure and protocols of network 26. Preferably, these modules also comprise appropriate traffic processing blocks, controllers and a protection interface, as

15

described above, in order to carry out the protection functions of the present invention.

Figs. 3A and 3B are block diagrams showing further details of gateway 30, useful in understanding a method of protection implemented by the gateway, in accordance with a preferred embodiment of the present invention. Fig. 3A represents normal operation of gateway 30, during which both modules 50 and 52 are operational. Under these conditions, RPR block 54a is preferably set to receive outgoing packets only on inner ring 32, while RPR block 54b is preferably set to receive outgoing packets only on outer ring 34 (although other divisions of traffic flow are also possible). As noted above, each of the RPR blocks captures packets addressed to only one of the MAC addresses of gateway 30, although both RPR blocks are configured in advance with both of the addresses. Both RPR blocks 54a and 54b can transmit incoming packets to nodes 28 on either of rings 32 and 34.

Fig. 3B illustrates the behavior of gateway 30 when a failure occurs in module 52, which is sensed by module 50 via protection interface 62 or by some other error indication. In accordance with the RPR protocol, as well as other bidirectional ring protocols known in the art, as soon as the nodes in network 22 become aware of the failure of module 52, the nodes on either side of the failed module wrap traffic away from it, or drop packets that were destined to pass through the failed module and steer traffic away from it. Thus, when wrapping is used, RPR block 54a of module 50 wraps traffic from inner ring 32 back onto outer ring 34, as shown in Fig. 3B, while node N4 (Fig. 1, not shown in Fig. 3B) wraps traffic from outer ring 34 onto inner ring 32. Alternatively, when

steering is used, block 54a drops traffic arriving on inner ring 32 that was destined to pass through module 52, and node N4 drops traffic arriving on outer ring 34 that was destined to pass through module 52. In either case, as long as the failure persists, packets reach RPR block 54a only on inner ring 32, at a data rate no greater than the bandwidth of the inner ring.

Meanwhile, controller 60a instructs RPR block 54a to begin capturing outgoing packets carrying the MAC address that is ordinarily assigned to module 52, rather than ignoring these packets as in normal operation. Since the RPR blocks are preconfigured with the MAC addresses of both of the modules in gateway 30, block 54a can implement this controller instruction automatically and very rapidly, typically within no more than 50 ms. This protection speed is comparable to that of standard unidirectional ring implementations, such as those based on SONET/SDH. At the same time, because of the inherent bandwidth limitation imposed by the wrapping mechanism shown in Fig. 3B or by the alternative steering mechanism, this rapid protection does not require that modules 50 and 52 have any more bandwidth than would be required of the active and standby modules in a unidirectional ring gateway.

Although preferred embodiments are described herein with reference to certain specific types of networks and protocols, and particularly to packet networks based on the RPR protocol, the principles of the present invention are similarly applicable in bidirectional ring networks and protocols of other types. It will thus be appreciated that the preferred embodiments described above are cited by way of example, and that the present

17

42165S2

invention is not limited to what has been particularly shown and described hereinabove. Rather, the scope of the present invention includes both combinations and subcombinations of the various features described hereinabove, as well as variations and modifications thereof which would occur to persons skilled in the art upon reading the foregoing description and which are not disclosed in the prior art.